

学校编码: 10384

分类号_____密级_____

学 号: 24320071151843

UDC _____

厦 门 大 学

硕 士 学 位 论 文

基于内容的视频检索技术研究

Research on Content-Based Video Retrieval

郑 银 环

指导教师姓名: 王 备 战 教授

专 业 名 称: 计算机软件与理论

论文提交日期: 2 0 1 0 年 5 月

论文答辩时间: 2 0 1 0 年 月

学位授予日期: 2 0 1 0 年 月

答辩委员会主席: _____

评 阅 人: _____

2010 年 5 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

厦门大学博硕士论文摘要库

摘要

随着计算机处理能力的日益增强,网络技术、多媒体技术的不断发展,视频数据、信息呈几何级数量增长。视频信息属于非结构化信息,具有复杂、庞大和冗余等特点,如何对其进行有效的管理和快速检索成为一个亟待解决的问题和研究的重点。因此,基于内容的视频检索技术应运而生。基于内容的视频检索直接对图像、视频等内容进行分析,抽取特征和语义,利用这些内容特征建立索引并进行检索,提供了一种符合人类认知规律的高效检索方法。作为一门涉及信息检索、图像处理、计算机视觉、人工智能等诸多领域的综合性技术,基于内容的视频检索已经成为国内外的研究热点。

论文主要工作表现在一下几个方面:

1. 深入研究了基于内容视频检索的研究现状及所存在的问题,针对视频数据缺乏清晰结构、非结构化等特点分析了现有的视频模型特点,结合 MPEG_7 标准改进了一种较为实用的视频分析模型。
2. 镜头检测是基于内容视频检索的基础,通过研究镜头变换的各种类型和表现,分析和比较了现有的非压缩域和压缩域镜头检测算法的优劣,详述了几种常用的镜头检测算法。
3. 关键帧获取是进行视频检索的必要步骤,通过比较基于镜头边缘、基于颜色特征、基于运动分析和基于聚类提取算法的优劣,提出一种基于主颜色直方图的关键帧提取方法。采用通用的视频间相似性度量算法和算法性能评判标准进行评价。
4. 设计实现了一个小型的视频检索原型系统用于测试和验证论文中提出的算法。根据用户提供的样例视频直接从视频源中检索出相似视频段,并对检索结果进行浏览。

关键词: 视频检索; MPEG_7; 内容检索

Abstract

With the development of computer processing capabilities, network technology, and multimedia technology, video data, information was of magnitude increase. Video information is unstructured information, which is complex, large and redundancy. How to effectively manage and quickly retrieve video from multimedia database is becoming urgent. Therefore, content-based video retrieval technology was proposed. Content-based video retrieval directly analyzes images and video content, feature extraction and semantic features which are used for index and video retrieval. As an integrated technology involved in information retrieval, image processing, computer vision, artificial intelligence and many other areas, content-based video retrieval has become a hotspot all over the world.

The dissertation makes a deep research on content-based video retrieval techniques, including the followings:

1. Firstly makes a deep research on the status and problems about content-based video retrieval. Because that video data is lack of clear description and unstructured, the dissertation proposes a more standard model for video analysis, combined with MPEG_7.
2. Shot detection is the basis for content-based video retrieval. The dissertation analyzes and compares the advantages and disadvantages of existing non-compressed domain and compressed domain shot detection algorithm, then introduces several common shot detection algorithm in detail.
3. Key frame for video retrieval is a necessary step, by comparing the merits of some key frame extraction, such as based on shot edge, based on color features, based on motion analysis and cluster extraction algorithm, the dissertation proposes a new key frame extraction that based on main color histogram. Evaluate the performance of the algorithm by adopting a common used similarity measure.
4. Finally Design and implement a small video retrieval prototype system for testing and verifying the algorithm proposed in the dissertation. It is used to retrieve similar videos from database and show the results to users.

Key Words: CBVR; MPEG-7; Content Retrieval.

目 录

第一章 绪论	1
1.1 选题背景及研究意义	1
1.2 研究现状及存在问题	4
1.2.1 CBVR 研究现状.....	4
1.2.2 CBVR 存在问题.....	6
1.3 论文主要研究内容和结构安排	8
1.3.1 论文主要研究内容.....	8
1.3.2 论文结构安排.....	9
第二章 基于内容视频检索体系及相关描述.....	10
2.1 CBVR 系统体系结构	10
2.2 数字视频结构	12
2.2.1 数字视频特征.....	12
2.2.2 数字视频描述.....	13
2.2.3 视频数据层次化结构.....	15
2.3 相似性度量	16
2.4 评判标准	18
2.5 小结.....	21
第三章 基于 MPEG-7 的视频检索及关键技术.....	22
3.1 MPEG-7 标准	22
3.1.1 MPEG-7 发展简介	22
3.1.2 MPEG-7 描述内容	22
3.1.2 MPEG-7 在基于内容的视频检索的应用	23
3.2 镜头边界检测	24
3.2.1 镜头转换.....	24
3.2.2 检测算法.....	25
3.3 关键帧提取	32
3.2.1 关键帧的特点.....	32
3.2.3 关键帧提取算法.....	33

3.4 视觉特征提取	35
3.4.1 颜色特征.....	35
3.4.2 形状特征.....	40
3.4.3 纹理特征.....	42
3.4.4 摄像机运动特征.....	44
3.5 小结.....	47
第四章 改进主颜色广义直方图的视频检索	48
4.1 算法研究基础	48
4.1.1 HSV 颜色空间.....	48
4.1.2 MPEG-7 描述子	49
4.2 算法思想	50
4.2.1 改进主颜色直方图关键帧提取.....	50
4.2.2 改进的初始主颜色确定.....	51
4.2.3 算法流程.....	53
4.3 实验结果及分析	54
4.3.1 系统实现.....	54
4.3.2 开发环境.....	56
4.3.3 算法测试结果.....	57
4.4 小结.....	58
第五章 总结和展望	59
5.1 总结.....	59
5.2 展望.....	60
参考文献	61
攻读硕士学位期间取得的主要科研成果	67
致谢.....	68

Contents

Chapter 1 Introduction	1
1.1 Background and Significance	1
1.2 Status and Problems	4
1.2.1 CBVR Research Status	4
1.2.2 CBVR Existing Problems	6
1.3 Dissertation Contents and Structure	8
1.3.1 Main Contents	8
1.3.2 Dissertation Structure	9
Chapter 2 Structure of Video Retrieval Based on Content and Description Involved	10
2.1 CBVR Retrieval Structure	10
2.2 Structure of video data	12
2.2.1 Character of video data	12
2.2.2 Description of video data	13
2.2.3 Layer Structure of video data	15
2.3 Similarity Measurement	16
2.4 Adjudicating Standards	18
2.5 Summary	21
Chapter 3 Video Retrieval Based on MPEG-7 and Key Skills	22
3.1 MPEG-7	22
3.1.1 MPEG-7 Introduction	22
3.1.2 MPEG-7 Description Contents	22
3.1.2 MPEG-7 Applying in Video Retrieval	23
3.2 Shot Boundary Detection	24
3.2.1 Shot Change	24
3.2.2 Detection Algorithm	25
3.3 Key Frame Extraction	32
3.2.1 Character of Key Frame	32
3.2.3 Key Frame Extraction Algorithm	33
3.4 Vision Character Extraction	35
3.4.1 Character of Colour	35
3.4.2 Character of Shape	40

3.4.3 Character of Texture.....	42
3.4.4 Character of Camera Moving.....	44
3.5 Summary.....	47
Chapter 4 Video Retrieval Based on Improved Main Colour Histogram	48
4.1 Basic Research on Algorithm.....	48
4.1.1 HSV Color Space	48
4.1.2 MPEG-7 Descriptor	49
4.2 Algorithm Process	50
4.2.1 Improved Key Frame Extraction Based on Main Colour Histogram ..	50
4.2.2 Improved Initial Main Color Determination.....	51
4.2.3 Algorithm Flow	53
4.3 Experiment Result and Analysis.....	54
4.3.1 System Implementation	54
4.3.2 Development Enviroment	56
4.3.3 Testin Result of Algorithm.....	57
4.5 Summary.....	58
Chapter 5 Conclusions and Future Work.....	59
5.1 Conclusions.....	59
5.2 Future Work.....	60
References.....	61
Publications	67
Acknowledgements.....	68

第一章 绪论

随着计算机处理能力的日益增强,网络、多媒体信息技术的迅猛发展,人类接触的信息量与日俱增。其中,多媒体信息(文字、图像、声音和视频等)是人们接触最广泛的一种信息资源^[1]。据调查统计,在人类获取的信息中,听觉信息约占 20%,视觉信息约占 60%,其他味觉、触觉、嗅觉总的信息加起来不超过 20%^[2]。视频片段是一种内容丰富、表达直观的多媒体信息,既有图像又有声音,兼有视觉和听觉信息的特征,具有文字和普通图像所无法比拟的丰富内涵和强大的描述能力,是人类获取信息的重要资源。视频信息是人类获取和传递信息的重要媒体介质和手段,它具有一般文字和普通图像所无法超越的重要作用。视频通过场景、镜头、关键帧能够描述更直观、生动、和具有丰富内涵的实体。

作为人们接触最广泛的信息资源之一,多媒体视频信息因其内容的丰富性、直观性以及巨大的信息量在多媒体信息系统、数字图书馆、数字博物馆、医学、卫星遥感、地质探测、科学考察、电影娱乐、视频点播、教育和在线信息服务、公共安全及国防等众多领域中扮演着越来越重要的角色。伴随着视频数据信息的爆炸性增长以及视频数据格式的非结构化特点,信息组织的无序性越来越突出、信息检索难度越来越大。如何有效地组织、管理大规模的视频数据库已成为当今计算机领域的重要课题;如何从浩瀚如海的视频信息中准确、高效地提取用户所需的资源也成为国内外研究的热点。

1.1 选题背景及研究意义

传统的视频检索技术是基于对视频内容描述的文本的检索,其基本原理如图 1-1 所示。首先给视频加上一个对其描述的文字或数字标签,然后在索引时对其标签进行索引。这种基于视频描述字符的匹配查询虽然简单且容易实现,但它却存在以下局限^[3]:

1. 对视频信息的文字标签索引缺乏统一的标准。文字描述是一种特定的抽象,如果描述的标准改变,则标签也必须重新制作才能适合新查询的要求。这些文字往往是观察者根据自己的理解加上去的,受主观因素影响

较大，不同的观察者或同一个观察者在不同的环境条件下对同一幅图可能给出不同的描述。没有统一的标准，常会出现自相矛盾的局面。

2. 文字标签不能完整概括视频信息内容。视频带给人的信息是多方面的，如颜色、模式、纹理、目标形状和分布位置等，用户需要访问的视频信息内容也是多方面的。简单的几个文字无法清楚地描述图像、视频等媒体所包含的丰富内容，因而，常常无法满足检索要求。
3. 难以满足实时信息检索要求。伴随着多媒体技术的发展，视频数据呈现爆炸式增长，视频数据库负荷日益加重。传统的手工标注要求在内容的浏览和查找中高度集中注意力，才能在视频、音频段中找到所要的内容。如此耗时耗力的查找方法明显无法适应网络中实时信息检索的要求。
4. 地球村概念的提出缩短了国家之间的距离，增进了不同国家间的交流合作，而文化、语言的差异成为沟通的自然屏障。对同一段视频的不同语言标注内容和习惯也直接影响了视频检索的精确度和精准度。
5. 原始视频由粗到细可以分为五个层次：视频流、场景、组、镜头和帧（静态图片）。帧是视频数据的最小单位，1 秒的视频约包含 24-30 帧，一段普通视频可能有几万到几十万帧。所以，直接对帧进行手工标注势必增加视频数据库的存储的负担，影响视频数据的检索速度。

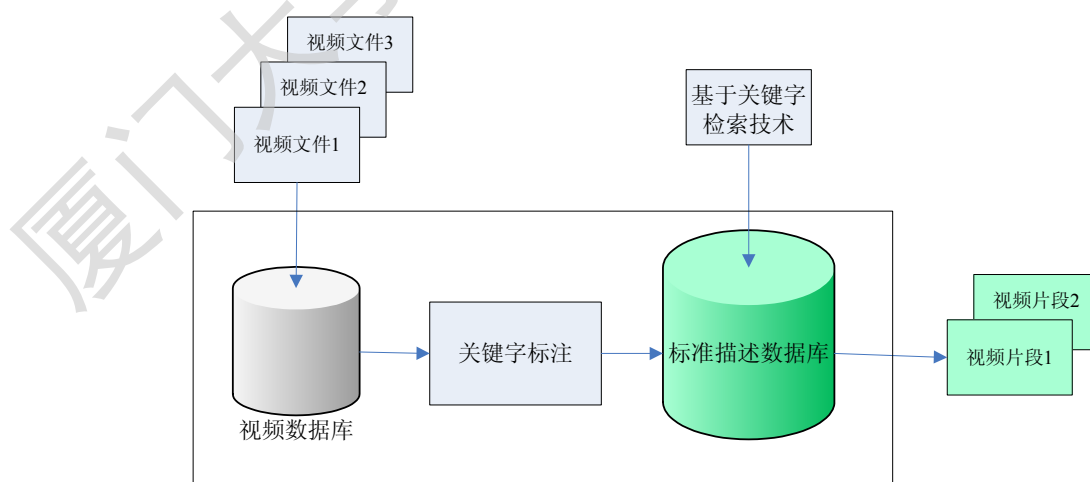


图 1-1 基于关键字的视频检索系统模型

为了避免上述基于标注的视频检索带来的问题，需要全面、客观地提取视频

内容，只有采用基于内容的检索方法才能提高视频检索准确率和效率，并在得到所需视频的基础上使视频信息得到有效的管理。基于内容的检索方法^[3]从新的角度管理多媒体信息，其中包括视频媒体的结构化组织和浏览；图像和音频信号的处理；对其他多媒体信息技术，如超媒体技术、虚拟现实技术、多媒体通信网络技术的支持等。基于内容的检索突破了传统的基于文本检索技术的局限，直接对图像、视频等内容分析，抽取对象的语义特征，包括：图像的颜色、纹理、形状、镜头、场景和镜头的运动等，利用这些内容特征建立索引，检索出具有与所提供的视频流、段相似的视频数据。基于内容的视频检索方法以图像处理、模式识别、计算机视觉和图像理解等领域知识为依托，融合了认知科学、人工智能、数据库管理系统、人机交互和信息检索等领域技术，从而设计出可靠、有效的检索算法。

与传统的信息检索方法相比，基于内容的检索方法有着无法比拟的优越性，其主要特点和优点体现在^[3]：

1. 对大型数据库的快速检索。多媒体数据库存储的视频数据具有数量巨大、内容丰富、种类繁多等特点，基于内容的检索方法能快速的从此类数据库中获得所要检索的视频段。
2. 相似性检索定义匹配。图像、视频一般为非结构化数据，数据之间的相似性度量比较模糊，检索时采用近似匹配，然后每次根据检索的结果逐步求精，不断减小检索范围，直至用户找到较理想的查询结果。
3. 从视频数据中直接提取信息线索，对视频数据进行低层特征和高层语义的分析。基于内容的视频检索技术突破了传统的基于文本标注检索的局限性，直接对视频、音频进行分析，从中抽取内容特征，然后利用这些内容特征建立索引并进行检索，大大降低了视频检索的主观影响，提高视频检索的准确率。
4. 满足用户多层次的检索要求。视频数据库由低到高可分为媒体库、特征库和知识库三种。媒体库一般为多媒体数据，如文本、图像、音频、视频等；特征库是包含用户输入的特征和预处理自动提取的内容特征；知识库包含领域知识和通用知识。针对用户的不同需求可以在不同层次检索，使检索过程更加人性化。

1.2 研究现状及存在问题

1.2.1 CBVR 研究现状

本世纪 90 年代后,快速发展的数据库管理和计算机视觉技术带来了 CBVR 的研究热潮。国际上,各大著名杂志纷纷设立专刊介绍该领域的研究现状与科研成果^[4],追踪 CBVR 相关技术的最新进展。国内外各大高校和科研机构也开始积极研究 CBVR,并发表了大量相关理论和技术应用的论文。

经过多年的理论研究,人们对 CBVR 在视频检索领域的重要性、有效性和优越性有了更深的认识,近年来大家开始把目光投注在实际应用上。各大科研、高校、商业机构不遗余力地投入人力物力研制成了一些系统^[5]。其中,采用全局匹配的有 IBM Almaden 研究中心公司的 QBIC (Query By Image Content) 系统,美国哥伦比亚大学电子工程系与电信研究中心图像和高级电视实验室联合研究开发的 VisualSeek 系统,美国伊利诺斯大学 Urbana-Champaign 分校开发的 MARS (Multimedia Analysis and Retrieval System) 系统,美国 MIT 多媒体实验室开发的 Photobook 系统。采用目标区域匹配的 UCSB 大学开发的 Netra 系统原型,斯坦福大学研究开发的 SIMPLIcity 系统,美国加利福尼亚大学伯克利分校计算机科学研究实验室开发的 BlobWorld,采用高层语义结合的由斯坦福大学研究开发的 a-LIP 和 ALIPR 系统。下面简单介绍几个比较具有代表性的系统^[6,7]:

1. QBIC 系统

QBIC^[8]系统是国际商用机器公司 IBM 的 IBM Almaden 研究中心于 20 世纪 90 年代开发的第一个商业化的基于内容的检索系统,它也是第一个功能齐全的视频数据库系统,是基于内容检索系统的典型代表,它的系统框架和结构是后来的图像检索系统的很好的参照,对视频数据库发展有较远的影响^[5]。QBIC 结合传统的文本关键字技术和 CBVR 技术,既支持基于示例图像的查询,也支持草图、轮廓和组合颜色、纹理和形状特征以及镜头和目标运动等信息的查询方式。QBIC 的高维特征索引结构的应用为以后的系统具有深远的影响。

2. VisualSeek 系统和 WebSeek

由哥伦比亚大学电子工程系与电信研究中心图像和高级电视实验室共同研

究的 VisualSeek 是一种基于视觉特性的搜索工具。VisualSeek 主要采用颜色集和基于小波变化的纹理特征进行查询,同时也支持它们之间空间关系的查询。另外,系统采用基于二叉树的索引算法加速检索过程。WebSeeK 作为 VisualSeek 的姐妹系统是一种在互联网上使用的基于内容的检索系统。它主要包含三个模块:图像、视频收集模块,主题分类和索引模块,搜索浏览和检索模块。和 QBIC 类似,WebSeeK 也支持基于关键字的查询,另外,它还支持在检出图像的基础上利用其可视属性做进一步的查询。

3. MARS 系统

MARS^[9]是美国 UIUC 开发的多媒体分析和检索系统。其特点是使用比较全面的底层特征,提供基于树结构的多特征的组合检索。由于采用多种特征描述与相似性度量方法,该系统提供较复杂的检索功能。

4. Photobook 系统

Photobook^[10]是美国麻省理工学院媒体研究室开发的支持相似性图像检索系统。该系统十分注重检索效率,着重解决预先设定的检索准则,在数据库中存储大量的信息以减少系统运行时的开销。它可以利用人脸、形状、纹理、等特征进行基于内容的检索。目前,Photobook 的人脸识别技术已经广泛运用在美国警察机关。

5. Netra 系统

Netra^[11]是 UCSB Alexandria Digital Library 开发的图像检索系统原型。该系统的一个主要特点是把对象分块,在各分块中使用颜色、纹理、形状、空间位置和运动特征等信息来检索视频库中具有和示例图像拥有相似分块的视频。系统主要研究基于 Gabor 过滤器的肺纹理分析、基于神经网络的图像分类构造和基于边流的区域分块。

6. TV-FI

Tsinghua Video Find It 简称 TV-FI 是国内较早开始研究内容检索的清华大学计算机系开发的视频节目管理系统。它提供多种访问视频数据库的模式,其中包括基于关键字的查询、基于示例的查询、根据视频结构的浏览以及按照用户预先定义类别浏览的方式。

7. ALIP 和 ALIPR

ALIP 和 ALIPR 系统^[12]是由 James Wang 和 Jia Li 在斯坦福大学设计雏形并在 Penn State 最终完成使用的一个系统。它是第一个基于高级语义的视频检索系统，根据视频数据给出语义级的标注分类。

此外，还有许多国家和地区所开发的类似的系统，如新加坡国立大学开发的 CORE 系统，哥伦比亚大学的 VideoQ 等。随着技术的不断发展和研究的深入，研究者通过开展一系列活动加强国际间的合作，促进相互间的学术交流以期达到优势互补，实现共同提高。

1.2.2 CBVR 存在问题

经过近二十年的研究和发展，CBVR 相关理论日趋成熟，应用也日益广泛，为视频检索的后续研究提供重要的基础。但视频信息的检索过程是相当复杂的，理论研究的最终目的是指导实践、应用于实际，这也是 CBVR 研究过程中遇到的关键问题之一。目前，CBVR 的研究焦点集中在对视频的分割方法、镜头的检测、关键帧的获取和帧的底层视觉特征如颜色、纹理、形状、空间信息和运动特征的提取上，由于视频结构、分割算法、镜头检测和特征描述、提取的复杂性，图像表层特征和人类认知的难以跨越的鸿沟，理论上仍然存在许多问题尚未解决。另外，由于 CBVR 技术的发展与计算机视觉、人工智能和机器学习等诸多领域的发展息息相关，目前这些领域的发展尚未成熟，在一定程度上限制了 CBVR 的发展。纵观 CBVR 的发展，对其的研究主要存在以下问题^[5,13]：

1. 视频数据的组织和管理

视频数据是一种非结构化数据，结构比较复杂。要对视频数据进行操作，首要解决的问题是如何对数据进行建模使视频数据在一定程度上结构化，即如何更有效、合理的组织视频数据，以便为后续的提取和分析、用户的浏览和检索提供便利。

2. 特征提取和描述

经过视频分割、镜头检测、关键帧提取等一系列处理后要进行图像特征的提取和描述，而颜色特征是大多数基于内容的检索系统的主要研究对象之一^[14]，

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库